

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Multicultural Assessment in Counseling and  
Clinical Psychology

Buros-Nebraska Series on Measurement and  
Testing

---

Spring 1996

## 1. Measurement, Testing, And Ethnic Bias: Can Solutions Be Found?

Stanley Sue

*University of California - Los Angeles*

Follow this and additional works at: <https://digitalcommons.unl.edu/burosbookmulticultural>

---

Sue, Stanley, "1. Measurement, Testing, And Ethnic Bias: Can Solutions Be Found?" (1996). *Multicultural Assessment in Counseling and Clinical Psychology*. 4.  
<https://digitalcommons.unl.edu/burosbookmulticultural/4>

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Multicultural Assessment in Counseling and Clinical Psychology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# MEASUREMENT, TESTING, AND ETHNIC BIAS: CAN SOLUTIONS BE FOUND?

Stanley Sue

*University of California, Los Angeles*

Assessment, evaluation, and diagnosis will gain increasing prominence as we head into the next century. Emphasis on managed care in the mental health system, well-being of individuals, job and work efficiency, personnel selection, upward promotions in one's career, admissions to institutions of higher education, etc., all require valid means of measurement and testing.

Several points are covered in this chapter. First, the assessment process involving ethnic minorities has many avenues by which bias can emerge. The biases can occur because of differences in culture or ethnicity as well as minority group status. Although culture has been defined in many different ways, it generally refers to the behavior patterns, symbols, institutions, values, and human products of a society (Banks, 1987). On the other hand, ethnicity can be used to describe a racial, national, or cultural group (Gordon, 1978). One's ethnicity typically conveys a social-psychological sense of "peoplehood" in which members of a group share a social and cultural heritage that is transmitted from one generation to another. Ethnic group members often feel an interdependence of fate with others in the group (Banks, 1987). In addition to culture and ethnicity,

members of ethnic minority groups also experience minority group status that involves a history of race or ethnic relations, a history that has affected interpersonal interactions, expectations, and performances. Thus to fully understand ethnic minority groups, their responses, and the assessment process, culture, ethnicity, and minority group status must be analyzed.

Second, concern with test and measurement bias is not simply a matter of being “politically correct” or of being perpetuated by ethnics who are disgruntled by their outcomes on various tests and measures. Bias does exist in many of our assessment instruments and procedures, and I shall try to demonstrate the range of biases using anecdotes and empirical evidence. Third, multiple steps should be taken to devise valid instruments and to understand the nature of cultural bias. Much of the research that will be cited involves Asian Americans; however, implications are drawn for ethnicity in general. Some anecdotal examples of sources of biases and consequences may more clearly indicate the importance of the issues to be presented.

### Some Examples of Sources of Bias and Their Consequences

1. In the development of the widely used *Diagnostic and Statistical Manual of Mental Disorders-III-R* (DSM-III-R) of the American Psychiatric Association (1987), Robert Spitzer contacted Arthur Kleinman, a prominent cross-cultural psychiatrist and anthropologist, for comments on cross-cultural issues. Kleinman (1991) wrote Spitzer a letter and was subsequently surprised to find that sections of his letter were compressed into two paragraphs of the introductory section of the DSM-III-R. He noted that considerations of the cultural limitations of the diagnostic system were too little, too late. Ethnicity and cross-cultural issues appeared more as an afterthought rather than a central variable. Fortunately, cross-cultural mental health researchers have been able to provide much more input into the recently published DSM-IV. Working groups were formed to offer recommendations concerning cross-cultural issues in diagnosis, and the DSM-IV has included discussions about cultural variations in symptoms of disorders as well as culture-bound syndromes. Although clearly an improvement over earlier versions, the DSM-IV still appears to lack a coherent approach to cross-cultural issues in psychopathology.

2. A concrete example of the consequences of inattention to ethnicity in assessment is demonstrated in the following case of a Chinese American psychiatric patient, David Tom, as noted in the *Seattle Times* (“The forgotten,” April 19, 1979):

The Cook County public guardian, Patrick T. Murphy, filed a \$5 million suit yesterday against the Illinois director of mental health and his predecessors, charging that they kept a Chinese immigrant in custody for 27 years mainly because the man could not speak English.

The federal-court suit charged that the Illinois Department of Mental Health had never treated the patient...for any mental disorders and had found a Chinese-speaking psychologist to talk to him only after 25 years.

The suit said that David, who is in his 50s, was put in Oak Forest Hospital, then known as Oak Forest Tuberculosis Hospital, in 1952.

He was transferred to a state mental hospital where doctors conceded they could not give him a mental exam because he spoke little English. But they diagnosed him as psychotic anyway.

The suit said that in 1971 a doctor who spoke no Chinese said David answered questions in an "incoherent and unintelligible manner."

It was charged also that David was quiet and caused little trouble but was placed in restraints sometimes because he would wander to a nearby ward that housed the only other Chinese-speaking patient. (p. A5)

(Incidentally, the patient did win his suit against the state of Illinois.) Although the patient may well have been psychotic, confidence in arriving at such a diagnosis would have been greater had a bilingual and bicultural mental health professional been available.

3. Korchin (1980) argues that in interpreting research findings on members of ethnic minority groups, there is often an implicit assumption that such findings must be compared with those on White Americans—the standard for comparisons. Under this assumption, ethnic minority group phenomena are not considered very important. For example, Korchin submitted to a major journal a coauthored paper assessing the determinants of personality competence among two groups of African American men—namely, those demonstrating exceptional competence and those demonstrating average competence. One of the journal reviewers indicated that the study was "grievously flawed" because there was no White control group. Korchin noted that the purpose of the study was to analyze within-group differences and not to compare African Americans and Whites. He then raised some interesting questions: "What would happen, might we suppose, if someone submitted a study identical in all respects except that all subjects were White? Would it be criticized



because it lacked a Black control group?" (p. 263). I am not implying that ethnic comparisons—something that we often do in research—are inappropriate. Rather, my contention is that we must interpret the research in an appropriate context and that ethnic group research is important in and of itself.

4. Several years ago, the American Psychological Association's Committee on Psychological Tests and Assessment was reviewing guidelines on assessment. In attempting to see that assessment procedures would not be culturally biased against ethnic minorities, the Committee dealt with a proposal indicating that if clinicians were not competent to conduct a psychological evaluation of an ethnic minority client—presumably because of cultural unfamiliarity—or if the assessment instrument was not validated on these clients, they should avoid making an assessment. One can imagine a similar proposal that if clinicians' competence with ethnic clients is in question, then they should not provide clinical services. Obviously, it would be inappropriate to subject ethnic minority clients to inadequate assessments or services. On the other hand, if the proposal had been adopted, the question would arise as to who would conduct assessments with ethnics. In other words, mental health professionals have the responsibility not only to decline from providing services when they are not qualified, but also to see that services are available to all. By simply admonishing clinicians to stay within their own areas of expertise, issues concerning accessibility of services, training of multicultural competencies in all clinicians, and development of cross-culturally valid assessment instruments are ignored.

These examples illustrate our neglect of cultural influences, assumptions about the standards of comparison by which to evaluate findings, and inability to foresee consequences of actions in trying to address ethnic minority issues. It is not surprising that in the case of ethnic minority populations, assessment has had a very controversial history. The controversy is over possible biases that occur when assessing the status of ethnic minority group individuals. These possible biases have been discussed over a diverse set of assessment tasks such as the ability to make valid assessment during clinical interviews, attempts to render a diagnosis, evaluations of client outcomes, estimating prevalence rates of mental disorders, use of personality inventories, use of cognitive and performance tests, etc. It is easy to understand the controversial nature of assessment among ethnic minority groups. Cultural considerations of minorities have not traditionally played a central role in guiding our assessment and evaluation efforts.

## DIFFICULTIES IN ASSESSMENT

In the assessment process, a number of problems can occur from a variety of sources in cross-cultural assessment. For example, Garcia (1981) argues that cross-cultural comparisons in IQ test performances fail to take into account possible cultural differences in motivation and task-relevant practice among test takers. Brislin (1993) takes issue with the equivalence of measures in cross-cultural assessment research: (a) translation equivalence, (b) conceptual equivalence, and (c) metric equivalence. Translation equivalence is a potential problem when questionnaires or instructions from one language group are used with another language group. It is based on the broader principle involving stimulus equivalence (e.g., whether a test item has the same meaning for different individuals). Translation equivalence exists when the descriptors and measures of psychological concepts can be translated well across languages. To test the translation equivalence of a measure that was developed in a particular culture, it is first translated by a bilingual expert to another language, then "back-translated" from the second language to the first by an independent bilingual translator. The two versions of the measure in the original language are then compared to discern which words or concepts seem to survive the translation procedures, with the assumption that the concepts that "survive" are translation equivalent. This procedure can be used to discover which psychological concepts appear to be culture-specific or culture-common.

Conceptual equivalence refers to the functional aspect of the construct that serves the same purpose in different cultures, although the specific behavior or thoughts used to measure the construct may be different. For example, one aspect of good decision making in the Western cultures may be typified by an ability to make a personal decision without being unduly influenced by others, whereas good decision making may be understood in Asian cultures as an ability to make a decision that is best for the group. These two different behaviors pertaining to making decisions are equivalent in that they comprise the very definition of the construct (good decision making) as used by individuals in the different cultures. Yet, the actual behaviors considered as good decision making are strikingly different.

Metric equivalence refers to the analysis of the same concept and the same measure across cultures, with the assumption that the scale of the measure can be directly compared across cultures. The assumption may be inaccurate. For example, a score of 100 on a certain scale or measure used with one population may not be equivalent to

a score of 100 on the same measure when used with a different population or when translated into another language. The lack of metric equivalence is especially apparent when cutoff scores are derived from one culture and then applied to another. Let us suppose that in the United States, a score exceeding 50 on a measure of depression is associated with severe clinical depression. This does not necessarily mean that in another country scores exceeding 50 on the measure are indicative of severe clinical depression. Norms for clinical depression as well as response sets to the measure may differ from culture to culture. These affect metric equivalence.

Potential problems in translation, conceptual, and metric equivalence have been sufficiently great that some researchers even go so far as to refrain from making any inference from the results of quantitative comparisons of a given measure between subjects from two different cultures (e.g., Hui, 1988). However, it is highly unlikely that comparisons between different cultural groups will discontinue, which makes it all the more important to test for, or develop, equivalency.

The person who uses professional judgement in assessment or evaluation is also subject to bias. This person and his or her evaluation process may be considered as a measurement "instrument." The reliability and validity of the counselor or clinician's assessment can be tested. The clinician is essentially an observer or a stimulus to the client and collects verbal and nonverbal data from clients. The clinician then performs a series of tasks such as making clinical judgments, inferences, and interpretations—all of which are subject to human biases, stereotyping, and faulty processing of information.

## EXISTENCE OF BIAS

Evidence has accumulated that suggests that assessments of individuals from culturally diverse populations are problematic (Jones & Thorne, 1987; Rogler, Malgady, & Rodriguez, 1989). Many investigators have suggested that cultural biases can affect therapists' interpretations of the psychological functioning of African Americans (Adebimpe, 1981; Mukherjee, Shukla, Woodle, Rosen, & Olarte, 1983; Neighbors, Jackson, Campbell, & Williams, 1989), American Indians (LaFromboise, 1988) Asian Americans (Li-Repac, 1980; Sue & Sue, 1987; Sue & Sue, 1991; Westermeyer, 1987), and Latinos (Good & Good, 1986; Lopez, 1989; Padilla & Salgado DeSnyder, 1985; Rogler et al., 1989). Because clinicians may not understand the cultural backgrounds or potential cultural response sets of ethnic minority clients, the validity of the clinical evaluations is open to questions.

In reviews of the literature, an overpathologizing bias (rating ethnic clients as being more disturbed than they actually are) was found by investigators who studied the validity of assessments of African American clients (Adebimpe, 1981; Neighbors et al., 1989). In one study, analysis of the records of 76 bipolar patients from different ethnic groups revealed that more than two-thirds of the clients had been previously diagnosed with schizophrenia (Mukherjee et al., 1983). The earlier diagnosis of schizophrenia was considered inaccurate because: (a) all patients demonstrated complete remission of psychotic symptoms without residual signs suggestive of schizophrenia; (b) the patients had been maintained on lithium, a drug commonly used to treat bipolar disorders, for an average of 3 years; and (c) not one patient's diagnosis was revised to schizophrenia. These data revealed that Latinos and African Americans were previously misdiagnosed with schizophrenia significantly more often than were White Americans.

It should be noted that overpathologizing is one direction of bias. Lopez (1989) has indicated that an underpathologizing bias (rating ethnic clients as being less disturbed than they actually are) can also occur. In his review of the literature, Lopez found that when instances of overpathologizing and underpathologizing are combined, substantial misdiagnosis of ethnics is found, and the evidence suggests that ethnic minority group individuals are more likely than are Whites to be assessed or diagnosed inaccurately.

Other studies have simply documented differences in evaluations as a function of ethnicity of therapists and clients. Li-Repac (1980) examined the influence of culture on the diagnostic approach of therapists. Five Chinese American and five White American male therapists rated the functioning of Chinese and White male clients during a videotaped interview. The results indicated that the ethnicity of both clients and therapists affected therapists' clinical judgments. Whereas White therapists rated Chinese American clients as anxious, awkward, confused, and nervous, Chinese therapists perceived the same clients as alert, ambitious, adaptable, honest, and friendly. White therapists rated White American clients as affectionate, adventurous, sincere, and easy-going, whereas Chinese therapists judged the same clients to be active, aggressive, rebellious, and outspoken. In addition, White therapists rated Chinese clients as more depressed, more inhibited, less socially poised, and having lower capacity for interpersonal relationships than did Chinese therapists. Chinese therapists rated White clients as more severely disturbed than did White therapists. These findings suggest that judgments about psy-

chological functioning depend at least in part on whether or not therapists are of the same ethnic background as their clients.

We (D. Fujino, G. Russell, S. Sue, M. Cheung, & L. Snowden) have recently completed a study examining the relationship between ethnic matches or mismatches between therapists and clients and therapists' evaluations of the initial level of functioning of clients. The study involved thousands of clients entering the Los Angeles County Mental Health System. Initial level of functioning was assessed using the Global Assessment Scale (GAS; Spitzer, Gibbon, & Endicott, 1985) in which clinicians provide a subject rating of the level of functioning of clients. Results indicated that ethnically matched therapists judged clients to have higher psychological functioning than did mismatched therapists. This effect held for ethnic clients (African, Asian, and Mexican Americans), but not for Whites. When the effects of other variables, such as age, gender, marital status, socioeconomic class, referral source, therapist's discipline, diagnosis, and gender match, were controlled, the effects of therapist-client ethnic matching were maintained for clients of African and Asian descent. Ethnic match was found to be a strong predictor of admission GAS scores, second only to diagnosis, a variable expected to be highly related to psychological functioning. The results are, indeed, provocative. Why do therapists who are of the same ethnicity as their clients evaluate the clients as being higher in level of functioning than do therapists who are ethnically dissimilar to their clients? We are not in a position to indicate the veridicality of the evaluations or to explain the findings because we could not randomly assign clients to therapists. Perhaps the clients who see ethnically similar therapists are simply less disturbed. Another possibility, consistent with Li-Repac's (1980) experimental study, is that therapists tend to rate ethnically similar clients as being less disturbed. In any event, much more research should be addressed to these possibilities. The main point is that clinicians or raters themselves are subject to biases.

Finally, what is it about ethnicity that may affect clinical judgments? Many researchers argue that the cultural orientation of therapists guides the diagnostic approach employed. If therapists fail to understand the cultural values, behaviors, assumptions about normality, and symptom expression of those from different cultures, the probability of making diagnostic and assessment errors is increased (Brislin, 1993; Good & Good, 1986; Rogler et al., 1989; Takeuchi & Speechley, 1989). For example, Asian Americans have been found to report somatic symptoms more than do White Americans (Sue & Morishima, 1982). It may be that such symptoms are more acceptable

in "face" oriented cultures, where having mental health disorders are quite stigmatizing and result in loss of face. Because people may learn to express distress in culturally acceptable ways, similar symptoms may hold different meanings in different cultures (Brislin, 1993). Thus, cultural modes of symptom expression can lead to misdiagnoses when clinicians do not understand the client's culture. Furthermore, it appears that the therapists' own sets of values and theoretical orientations influence their evaluations of client behavior (Rogler et al., 1989). For example, the Chinese and White clinicians in Li-Repac's study (1980) made different evaluations about the functioning of clients even though they viewed the same videotaped interviews.

Obviously, cultural factors may bias assessment and confound our interpretations. However, it is also possible that observed assessment differences between culturally different groups are real. For example, in a study by Keefe, Sue, Enomoto, Durvasula, and Chao (in press), the MMPI-2 performances were examined of Asian American and White students. Additionally, Asian Americans completed the Suinn-Lew Self-Identity Acculturation Scale (SL-ASIA; Suinn, Rickard-Figueroa, Lew, & Vigil, 1987). We divided the Asian Americans into those who were more acculturated and those who were less acculturated. The findings indicated that less acculturated Asian American students showed greater elevation on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Hathaway, McKinley, & Butcher, 1989) profile than did more acculturated Asian American students or White students. Furthermore, more acculturated Asian American students had greater elevations than did their White counterparts. On individual MMPI-2 scales where differences were found, scale elevations were largely ordered in the following manner: Less acculturated Asian Americans > acculturated Asian Americans > Whites. (On the validity scales, the three groups did not significantly differ, except on the F Scale in which less acculturated Asians were higher than Whites.) The results can be interpreted in at least two ways. First, the results may suggest that Asian American students had more psychopathology than did Whites. Moreover, less acculturated Asian Americans were particularly high in disturbance. It could be argued that such findings reflect the fact that Asian Americans are under greater stress because of culture conflict, adjustment to a new environment, language problems, minority group status, and so forth. This may be especially true of the unacculturated.

Second, the ethnic differences may result from the metric nonequivalence of the scores or from response sets that vary from one cultural group to another. Response sets include acquiescence (e.g.,

tendency to agree with statements) and social desirability (i.e., answering in ways that are intended to create an appropriate or good impression on others). Thus, Asian Americans may not actually be more disturbed; rather, the assessment tool and the inferences drawn may not be equally valid for different groups. If this is the case, then the personality inventory must somehow be corrected or modified in order to provide an accurate assessment of Asian Americans. Without examining culture and cultural bias, finding an explanation for the results is problematic.

It should be noted that studies of bias are difficult to conduct in the mental health field because we often have no absolute criteria by which to unequivocally judge the accuracy of evaluations. In Li-Repac's experimental study (1980), evaluations of clients varied as a function of ethnicity of therapists and clients. However, this question remains unanswered: Which ethnic group therapists were more accurate in their judgements?

There are other means of assessing bias in tests, and two of the most popular include factor analysis and regression analysis. If the factor structures are different for different populations, the instrument is not tapping into the same phenomena for the populations. Regression analysis can be applied to see if the tests make similar, and similarly accurate, predictions between the tests and a criterion measure. If, for example, regression slopes for a test or evaluation procedure and a criterion differ for different groups, test bias exists. Such studies require that we have fairly clear-cut criteria on which to judge the adequacy of predictors. Although some researchers (Kaplan & Saccuzzo, 1982) believe that slope bias for ethnic minority groups has rarely been demonstrated in empirical studies, we found convincing evidence for slope bias in the case of Asian Americans. Let me now turn to some of our research on educational achievements among Asian Americans (Sue & Abe, 1988) in order to demonstrate some major biases in assessment.

## PREDICTORS OF ACADEMIC ACHIEVEMENTS

In response to concerns over university admissions policies and criteria for admitting students, the University of California system collaborated with the College Board to investigate the validity of various predictors of academic achievement for Asian American students. Examined were Asian American students who enrolled as freshman in any of the eight University of California campuses during fall 1984. The campuses included Berkeley, Davis, Irvine, Los Angeles, Riverside, San Diego, Santa Barbara, and Santa Cruz. The pur-



pose of the study was to determine how well certain variables such as high school grades and SAT scores predicted academic performance during the freshman year. The study was unique in that no other validity investigation had examined differences among various Asian American subgroups on these factors, nor had any other study reported on as many Asian American students.

In terms of the design, we examined the records of the 4,113 Asian domestic (nonforeign) freshman students who enrolled in any of the eight campuses and compared them with those of 1,000 randomly selected White students. Males constituted about 50% of the Asian Americans, whereas 49% of the White sample were males. The Asian American student numbers were, in descending order: Chinese 1,470, Filipinos 712, Japanese 643, Koreans 575, Other Asian Americans or those not members of the specific groups listed in this study 525, and Asian Indians/Pakistanis 170.

The criterion variable was the university freshman grade point average (GPA), which was the average of all grades received by a student during the academic year. Different predictor variables were used for the GPA. I shall only report on high school grade point average (HSGPA) calculated from courses and Scholastic Aptitude Test-Verbal and Scholastic Aptitude Test-Mathematics scores. HSGPA, SAT-V score, and SAT-M score were used as predictors of university grades. This set of variables has been widely employed in making admissions decisions and was of primary interest in this study. Regression analyses were performed for each Asian American group, all Asian American students combined, and Whites. Analyses were also made for all Asian Americans and Whites, according to sex and academic majors.

## General Results

Let me briefly present the results. First, Asian American students were found to have superior high school grades compared to Whites. Considerable within group differences were found with Asian Indians/Pakistanis having the highest and Filipinos having the lowest mean HSGPA. With the exception of the Filipinos, all the Asian American subgroups exceeded the average HSGPA of Whites. Regardless of ethnicity, females had higher HSGPAs than did males. Second, consistent with previous studies, Asian Americans achieved higher average SAT-M scores than did Whites; they received lower average scores than did Whites on the SAT-V sections. For both Asian Americans and Whites, males had higher SAT-V and SAT-M scores than did females. Thus, although females exceeded males in high



school grades, their average SAT scores, particularly on the mathematical portion, were lower than those of males. Large differences in SAT performances were found among the Asian American subgroups, with Asian Indians/Pakistanis having the highest SAT-V score, and Koreans having the lowest. On the SAT-M test, the Chinese scored the highest and Filipinos scored the lowest. Third, the university grade point averages for Asian American and White students were very similar. Whereas Asian American males and females were highly similar in GPA, White females tended to achieve higher grades than White males did. Within the Asian American student group, considerable ethnic differences in university GPA were found. In descending order, the mean GPAs for the groups were Chinese, Asian Indians/Pakistanis, Other Asians, Japanese, Koreans, and Filipinos.

### High School Grades and SAT Scores as Predictors of University Grades

The most interesting results concern the ability of high school grades and SAT scores to predict university grades. Multiple correlations were used to note the contributions of the predictors to university grades. Let me summarize the findings. Whereas HSGPA made the largest contribution in the prediction of university grades for both Asian Americans and Whites, considerable differences were found in the contributions made by SAT performances. For Asian Americans the SAT-M score contributed more to the prediction of university grades than did SAT-V. For Whites the situation was reversed; SAT-V made a larger contribution to university grades than did SAT-M. Dividing the students by ethnicity and sex did not alter the findings. Some marked differences emerged when the various Asian American groups were compared. We also tried to analyze the ability of the SAT to predict grades within academic majors in order to find out if the superiority of math over verbal skills was specific to those students in quantitative fields. The overall results generally persisted in that regardless of majors, SAT-M tended to be a better predictor of grades for Asians than for Whites.

Another way of comparing ethnic differences in predictors of academic achievement is to examine the possible prediction bias that occurs when the regression equation derived from one group is applied to the other. In other words, is the regression equation generated by Whites accurate in predicting the performances of Asian American students? We wanted to use Whites because this population, rather than ethnic minority groups, is likely to be the standard of comparison. To derive the White regression equation, a standard

least squares regression was performed. By entering into this equation the scores received by Asian American students on the predictor variables, we could compare the grades predicted by the White regression equation with those that were actually received by Asian American students. Asian Americans received actual grades that were .02 higher than the predicted grades. Thus, using the White regression equation for Asian Americans placed Asian Americans at a slight disadvantage. Some substantial differences occurred, however, when the prediction bias was examined for specific groups. The White regression equation severely underpredicted the performances of Chinese and Other Asian American students. For example, Chinese students were predicted to have a grade point average of 2.77, when they actually had an average of 2.89. Although GPA differences of .10 or .20 may seem slight, they are very important not only to the student but also to graduate programs which must often make difficult decisions about the students to admit. Serious overprediction occurred for Filipinos and Japanese. This means that the White regression equation was biased in either direction, depending on the particular Asian American group. Obviously, if the regression equation derived from the Chinese sample is used for other Asian groups (or for Whites), we would also find prediction bias. It is not surprising that the application of one sample's prediction equation to another sample results in decreased accuracy for the other sample.

The purpose of the study was to examine the validity of predictors of first-year university grades for Asian American and White students. The findings can be summarized as follows: (a) High school grades and SAT can, to a moderate degree, predict university freshman grades of Asian American and White students. (b) Consistent with findings from other studies, the best single predictor for all students was the high school grade point average. (c) For Asian American but not for White students, mathematics scores or quantitative skills are a better predictor of university grades than are verbal scores. This ethnic difference persisted even across academic majors declared by students. (d) No major sex differences emerged to contradict the *overall* ethnic differences that were found. (e) The various Asian American groups showed interethnic differences in the proportional contributions of high school grades and SAT scores in the prediction of university grades. (f) The White regression equation underpredicted or overpredicted the performances of Asian Americans, depending on the particular group.

The strength of this study was the inclusion of a large Asian American student sample broken down by particular ethnicity. How-

ever, there are some important limitations to consider. For example, it was not possible to examine other important variables such as the socioeconomic class of the students, which may substantially influence the validity of predictors. Also, the sole criterion of overall achievement was first-year university grades. Other criteria should be used, such as grades in certain courses, grades for more than just the freshman year, or nonacademic indices of achievement. These limitations suggest that further research is needed in order for us to understand the theoretical and policy-related issues involved in the academic achievement of Asian American students.

This study demonstrates that in something as important as prediction of university grades, substantial ethnic differences exist in predictor-criterion relationships. The use of a regression found for one ethnic group may present a seriously biased picture for members of another ethnic group. The problem is that in practice a single prediction equation may be used, based on the dominant or majority group, which then reduces the validity of the prediction for members of minority groups. Assuming that one major goal of admissions criteria is to enroll the best students, it is interesting to note that I know of no university that has tried to use group specific regression equations in the selection of Asian American students. I am not arguing that English verbal skills are unimportant. Rather, if we want to select the best students—at least in terms of freshman grades—then mathematics scores should be weighed more heavily than verbal skills among many Asian American groups.

## ADDRESSING ASSESSMENT BIAS

Given that tests and measurements of ethnic minority group populations are problematic and subject to bias, the question arises regarding what can be done. Several tasks should be considered. Let me briefly outline six major tasks, discussing in more detail the last three in which my colleagues and I have been involved.

### Devise New Tests and Measures

New psychological tests and measures that are appropriate for ethnic minority populations need to be developed. I can think of three areas where new tests and measures would be very helpful. First, alternative measures for assessing attitudes, personality, and behaviors are a potentially fruitful area of investigation. Two decades ago, Robert Williams (1974) attempted to establish the Black Intelligence Test of Cultural Homogeneity, a intelligence test that is heavily loaded on items that are more specific and familiar to African Ameri-

cans than to Whites. Although the validity of the test for predicting intellectual functioning has been controversial, Williams' work highlighted the importance of culture in influencing performance in at least some of the items typically used in IQ tests. Mercer (Mercer & Lewis, 1979) has also established the System of Multicultural Pluralistic Assessment, which is another attempt to take into consideration cultural elements in intellectual performance. Such efforts should continue because they bring into the forefront issues concerning the nature of what we examine (e.g., what is IQ?) and the impact of culture in the tests. New tests should be devised as alternatives to what is available.

Second, assessment of concepts that are pertinent to cross-cultural concerns are also important to assess. For example, researchers have been trying to develop means of measuring acculturation (Cuellar, Harris, & Jasso, 1980; Sodowsky & Plake, 1991; Suinn et al., 1987), ethnic or racial identity (Helms, 1990; Helms & Carter, 1991; Mendoza, 1989; Phinney, 1992), or multicultural competence and the elements comprising competence in counseling (see Ottavi, Pope-Davis, & Dings, 1994; Ponterotto & Casas, 1991; Sodowsky, Taffe, Gutkin, & Wise, 1994). The research is significant because the findings provide important knowledge of the similarities and differences within and between ethnic groups, social development associated with cultural practices, self-esteem and well-being, and cross-cultural competencies. In these areas, cross-cultural and ethnic minority researchers can provide special expertise.

Third, we should develop new measures that evaluate important values or traits that have salience especially for ethnics. As an illustration, let us examine personality assessment. In the United States, researchers have unearthed five orthogonal personality factors, called the "Big Five" (Goldberg, 1981), that include characteristics such as agreeableness, conscientiousness, and emotional stability. It is likely that these five factors have importance to a greater or lesser degree across different cultures (Yang & Bond, 1990). Nevertheless, the question remains of whether for certain ethnics other characteristics may be more salient or important than the Big Five as personality dimensions. One of my colleagues, Nolan Zane, is trying to address this issue with Asian Americans. He believes that one significant personality attribute that affects interpersonal interactions is "face." Loss of face (defined as the threat or loss of one's social integrity) has been identified as a key and often dominant interpersonal dynamic in Asian social relations, particularly when the relationship involves help-seeking issues among Asian and White students. Many indi-

viduals fear the loss of face or their social integrity, particularly Asian Americans who come from face cultures. Zane (1991) has developed a loss of face measure (LOF). The 21-item measure reflects four face-threatening areas involving social status, ethical behavior, social propriety, and self-discipline. Preliminary findings indicate that the measure has good reliability and validity. It correlated positively with other-directedness, self-consciousness, and social anxiety and negatively with extraversion and acculturation level of Asian Americans. Asian Americans also score higher on the measure than do Whites. LOF appears to be able to predict, independently of social desirability, certain behaviors such as assertiveness and help-seeking behaviors. Zane suggests that certain personal constructs may be more culturally salient for some groups than others.

### Evaluate Tests and Revise to Make Them Cross-Culturally Valid

Most research on assessment with ethnic minority groups has examined the use of existing instruments. Many studies have tried to determine the validity of instruments, derived in the West, when used with members of ethnic minority groups or cross national populations. Intelligence tests (e.g., the Wechsler Adult Intelligence Scale [WAIS]), personality inventories (e.g., MMPI-2), and survey instruments (e.g., Diagnostic Interview Schedule) have been employed in the study of ethnic minorities or cross-national groups. Rogler, Malgady, and Rodriguez (1989) indicate that common problems include not only translation equivalence and item familiarity but also assumptions concerning the meaning of responses to items. With respect to meaning of responses, they note that in Puerto Rican culture spiritualism is practiced and that answering affirmatively to MMPI items, such as "Evil spirits possess me at times," may not be indicative of pathology. Under such circumstances, the instruments can be modified in order to enhance their validity or local norms can be established with different populations. Such efforts are important in that they provide a standard by which to compare different groups and yield insights into what aspects or items of a measure are cross-culturally appropriate or inappropriate and what modifications may be necessary in order to strengthen validity and to more accurately interpret test results.

### Advocate for Cross-Cultural Considerations and Policies

We have certain roles to perform as assessment researchers and practitioners. Involvement in our professions should also include participation in the formulation of policies and practices, if we are to

have an impact on assessment. We should caution others about the difficulties in conducting assessments of members of ethnic minority groups and advocate for the integration of cross-cultural considerations in research, theory, and assessment practice. After all, psychology and the social sciences involve the study of human beings and not of a particular group. In order to affect assessment policies and practices, cross-cultural assessment experts should be included in all boards, committees, policy-making groups in organizations such as the American Psychological Association, American Psychiatric Association, American Educational Research Association, and American Evaluation Association, as well as in state and local governmental agencies that deal with assessment. They should also have strong input into all policies concerning the use of assessment tools and the appropriateness of assessment procedures.

### Adopt New Assessment Research Paradigms

A variety of research strategies have been used in cross-cultural psychology. The strategies can be classified as (a) point research, (b) linear research, and (c) parallel research (Sue & Sue, 1987; Zane & Sue, 1986). Each progressively helps to uncover the meaning of assessment in cross-cultural comparisons.

*Point research.* Point research simply compares the performance of one cultural group with another. It is the most frequently used cross-cultural approach. In most cases, an assessment instrument developed in one culture is used in another culture. Often, the scores on the instruments are compared between the different cultures and interpreted from the norms developed from one culture. Because of the relatively long history of psychology in Western societies, many of the instruments are of American or Western European origin, frequently requiring language translations for use with non-English-speaking groups. For example, we (Chu, Lubin, & Sue, 1984) have translated the Depression Adjective Checklist and studied the reliability and validity of the instrument for Chinese in Taiwan. The use of measures developed in one culture and applied in another culture runs the risk of perpetuating an imposed emic in assessment. That is, taking an emic (culturally specific) assessment scale and using it as if it were etic (universally applicable) in nature can be a serious problem. Researchers are increasingly aware of potential problems caused by an imposed emic, but for many cross-cultural investigators, more safeguards should be used.

As mentioned earlier, several assumptions underlie the development of a cross-cultural measure. It is assumed that the concept as

measured by the instrument exists in both cultures, that the concept is equivalently operationalized, and that there is scalar or metric equivalence of the instrument. Violation of these assumptions frequently occurs in cross-cultural research (Hui & Triandis, 1985). Other cultures may not have the concept under investigation or may define it differently (Dohrenwend & Dohrenwend, 1969). In using the Beck Depression Inventory among Vietnamese populations, Kinzie, Manson, Vinh, Tolan, Anh, and Pho (1982) found that the Beck Depression Inventory was not reliable or valid in the diagnosis of depression. This may be the result of cultural differences in conceptualization of depression or in symptom manifestations of the same disorder. Investigators (Kleinman, 1977; Sue, Wagner, Ja, Margullis, & Lew, 1976; White, 1984) have found that some constructs derived from the Western perspective are conceptualized differently or do not exist in other cultures. The difficulty involved in translating words used on assessment devices may be an indication that the concepts may not be equivalent. In view of these potential problems, the mere fact that different cultural groups exhibit differences on a particular assessment measure suggests that the groups *may* differ. Point research should be supplemented by linear research in order to more firmly establish that the differences found in point research are real.

*Linear and multimethod models.* In trying to validate measures, researchers often see if the measure relates well to other measures or indices of the construct under investigation or if the measure is a good predictor of the phenomenon being studied. For example, if an intelligence or cognitive measure, which was originally developed and validated in the United States, is a valid indicator of intellectual functioning in Japan, we would expect the measure to: (a) correlate well with other measures of intelligence among Japanese, and (b) predict the future performance of Japanese, for instance, in academic performance. If the measure shows little concurrent or predictive validity among Japanese, then it may be poorly suited for cross-cultural use.

Linear research is intended to examine the validity of an instrument. Whereas point research establishes that two cultural groups differ on a measure, linear research tries to establish whether the differences are real or an emic artifact of the measure. A series of studies using different measures of a construct can be used with two or more culturally distinct groups, or different measures can be used in a single study. For example, Sue, Ino, and Sue (1983) wanted to study assertiveness among Asian American and Whites and used a multimethod strategy. In this study, individuals were administered



paper-and-pencil tests, typically used in studies of White Americans, as well as behavioral measures of assertiveness. The self-report, paper-and-pencil measure supported the notion that Asian Americans are less assertive than their White counterparts. However, no overall differences on behavioral measures were found. The finding that Asian Americans could behave as assertively as their comparison group raises questions about the validity of the paper-and-pencil measure.

Another example of the linear approach can be seen in the series of studies reported by Dohrenwend and Dohrenwend (1969). The investigators wanted to study the prevalence of psychopathology among different ethnic groups in the United States. The strategy employed was based on point research in which different cultural groups are compared on a measure. After administering the Midtown 22-item symptom questionnaire, they did find ethnic differences: Puerto Ricans scored higher in psychological disturbance than did Jewish, Irish, or Black respondents in New York City. But how did they know if the Puerto Ricans were actually more disturbed or if the findings were simply an artifact of the measure? That is, the findings may simply indicate that the instrument failed to have cross-cultural validity. Fortunately, the Dohrenwends then adopted a linear research strategy to test whether the higher score among Puerto Ricans indicated higher actual rates of disorders. In a subsequent study, they matched patients from each ethnic group in terms of psychiatric disorders and administered the same questionnaires as before (i.e., the Midtown 22-item symptom questionnaire). Because patients were matched on type and presumably severity of disorders, one would expect no differences in symptom scores. However, Puerto Ricans again scored higher than the other groups. Dohrenwend and Dohrenwend argued that the higher scores for Puerto Ricans probably reflected a response set or a cultural means of expressing distress on the questionnaire rather than actual rates of disturbance. Their conclusions were based on a series of studies trying to ferret out cultural factors from actual psychopathology in the analysis of the measure.

*Parallel Research.* Unlike the point approach in which differences between ethnic groups are examined on a particular measure, and the linear approach in which researchers try to establish if observed group differences are real, the parallel research strategy is intended to explain any real differences that are found. Explanations for behaviors often differ from one culture to another. In parallel research, the task is to develop means of conceptualizing the behavioral phenomena from the different cultures in question. A parallel design is



essentially two linear approaches, each based upon its own cultural viewpoint. Previously, I discussed the issue of decision making. If we constructed a Western measure of decision making, individuals from nonWestern cultures might reliably differ on the measure and appear to have deficits. Only by adopting each cultural explanation can we truly understand that in some Western cultures good decision making involves making independent judgments whereas in some Eastern cultures good decision making is associated with doing what is best for the group. The advantage of this design is that the framework or perspective from one cultural group is not imposed on another. In this way, similarities and differences of the construct or concept under investigation can be determined. This can be illustrated in research on depression among Asian Americans.

Clinical folklore among researchers and practitioners suggests that Asian Americans may express depressive symptoms differently from White Americans. Asians often seem to manifest somatic symptoms rather than strict depressive symptomatology, such as self-reports of sadness or dejection (Sue & Morishima, 1982). Thus, it is unclear if measures of depression, used in the United States, can be appropriately applied to Asian Americans. Kleinman (1977) believes that depression is conceptualized differently by certain Asian groups and that attempts to study depression in other cultures by using Western-derived criteria such as those listed in the *Diagnostic and Statistical Manual* of the American Psychiatric Association may be misleading. Given the uncertain validity of depression measures and possible cultural differences in the expression and conceptualization of depression, Kinzie et al. (1982) adopted a research approach much like the parallel research strategy described above, in developing a depression scale for Vietnamese. In the United States, relatively much research has been conducted on the assessment and measurement of depression, and the symptoms and syndromes associated with depression among White Americans have been identified. However, this is not the case with Asian American groups such as Vietnamese Americans. Therefore, a parallel strategy would entail the development and validation of a depression measure based on indigenous (i.e., Vietnamese) conceptualizations of the disorder and the analysis of the reasons why White and Vietnamese Americans may differ in the disorder or its manifestations.

To begin the task, four bilingual mental health workers, who worked independently, generated a list of Vietnamese words that were related to depression in the areas of thinking, feeling, and behavior (items associated with DSM III criteria for depression were

given consideration). The adjectives were then compared and revised in terms of lexicon and grammar. Interestingly, the investigators used a 3-point rather than 5-point Likert scale because they found that Vietnamese felt that five rating levels would not be sensible to their cultural group. The items were then translated into English and back-translated into Vietnamese to check semantic integrity. They were then administered to a small group of Vietnamese as a pretest to test for sensibility and appropriateness. Any items that needed explanation and that proved to be inappropriate were revised. To validate the scale, scores on the scale from a depressed Vietnamese clinic sample were compared with those from a demographically matched community sample of Vietnamese adults. The comparisons showed that the depressed clinic sample and the control sample differed significantly on the majority of items (27 of 45). Surprisingly, only 4 out of the 27 items that were statistically significant between the depressed and control groups were similar to those in the DSM III (these were psychophysiological symptoms). The other 23 were from Vietnamese descriptions of cognitive, affective, and somatic indicators of depression.

The symptoms of depression that were common in Vietnamese and Western cultures were primarily somatic, or psychophysiological, in nature: Poor appetite, headache, poor concentration, and exhaustion. However, those items indicative of moods, such as "sad and bothered," "low spirited and bored," and "downhearted and low spirited," were more difficult to interpret. These phrases were not overlapping (i.e., not much commonality was found in Vietnamese and Western cultures). About two-thirds of the items were unrelated to items often associated with the Western conception of depression, including, "being angry," "feeling shameful and dishonored (not guilt)," "feeling desperate," and "having a feeling of going crazy." The results demonstrated that conceptualizations of disorders do differ among different cultural groups. Kinzie and his colleagues reported difficulty in translating many of the Vietnamese concepts and stated that "the lack of one-to-one correspondence also suggests that the meanings of particular Vietnamese thoughts, feelings, and behaviors may be different from our own, are implicated, and cannot be adequately conceptualized apart from a broader semantic network" (p. 1280). In summary, the results of the work by Kinzie and his associates indicate that there is some overlap in the symptoms reported by both Western and Vietnamese cultures. However, the nature of differences may indicate that the symptoms do not reflect the same construct. Responses to assessment measures may then vary according to culture and can be explained by different cultural construals of depression.

## Study the Nature of Bias

One research area that has been largely ignored in cross-cultural assessment is that of bias. Although many scholars have discussed the nature of bias and have offered conceptual analyses of it, we lack empirical research into the origins of bias. Let me explain our research in this area. Our research program is intended to study the response sets or cultural dimensions that may operate when Asian Americans are administered measures of psychopathology developed in Western societies. The ultimate goal of the research is to understand cultural processes that influence responses to assessment instruments and, with this understanding, to increase the validity of the instruments for Asian American populations. The research was undertaken for several reasons. First, current assessment tools that are widely used in the United States have been criticized for not taking into account cultural factors that may bias evaluations for ethnic populations in general and Asian Americans in particular. Second, although research and clinical assessment instruments are continually being revised and modified in order to achieve greater reliability and validity, the adequacy of the instruments is rarely examined for Asian American populations because they are relatively small in numbers. When validation studies for Asian Americans are conducted, they tend to occur many years after an assessment tool is developed. By that time, new instruments have been devised and Asian American researchers are then studying the validity of an "old" instrument. Third, validation studies simply tell us whether or not an instrument is appropriate for a given population. If the instrument is inappropriate, the reasons and underlying processes for the lack of validity are a matter of speculation. Finally, although the obvious solution would be to design a valid assessment tool specifically for Asian Americans, there are many practical problems in devising a culture-specific measure, and such a measure would not allow comparisons to be made with non-Asian populations. A culture-specific measure may be appropriate and helpful in some situations, as mentioned earlier. However, our research plan is to gain insight into the processes underlying Asian American responses to assessment instruments—processes and principles that may have generality across different assessment tools.

The proposed pilot research is important in discovering sources of bias and means of correcting the bias. The findings can be used to evaluate all inventories, because underlying dimensions or processes are identified. In turn, the validity of measures for clinical and epidemiological use with Asian American populations will improve,

because the identified biases can be controlled. This means that researchers can *continue* to use existing, mainstream, and traditional measures with Asian Americans. Rather than to abolish such measures or construct some measures that are specific to Asian Americans, one simply needs to control for identified biases in existing instruments. For example, the development of a social desirability measure has enabled researchers to control for socially desirable responses on existing personality and psychopathology measures.

At a more basic or theoretical level, the research can lead to a greater understanding of the factors that influence responses on measures of psychopathology, especially cultural-based ones. In the past, researchers have suggested that factors such as cultural differences in shame and stigma, response sets such as social desirability, concepts of mental illness, etc., have hindered an accurate assessment of various ethnic minority groups including Asian Americans. Our task has been to see if ethnic differences in responding can be predicted by cultural response sets (or cultural dimensions). We want to see if certain cultural dimensions (such as shame and stigma, tolerance for symptoms, and cultural familiarity with symptoms), which some investigators have proposed as being important for Asian Americans, can predict performances on certain measures. Our study has several steps:

1. *Identify cultural dimensions that differentiate responses of Asian Americans and Whites on self-reported measures of psychopathology.* Researchers have often speculated that Asian Americans and Whites differ in cultural variables (e.g., shame and stigma and self-disclosure) or response sets (e.g., social desirability) that may influence responses to personality or psychiatric inventories. The project empirically examines how Asians and Whites differ on their evaluations of individual questionnaire items in terms of stigma, cultural familiarity, etc. Using this method, those items or sets of items on questionnaires that are likely to demonstrate ethnic differences on the basis of cultural response sets can be identified. We can also determine which instruments are heavily loaded on cultural response sets and likely to give biased findings.
2. *Use the identified dimensions to construct scales in a major study that can be used to control for bias in order to increase cross-cultural validity.* Once dimensions have been identified as being important, scales can be developed to represent the dimensions. The scales can then be used to control for cultural bias. For example, if Asian Americans tend to underreport symptoms that arouse feelings of shame, and a particular questionnaire is

heavily loaded on items involving shame, a shame scale can be constructed and used to control for the underreporting.

The proposed research program investigates the effects of cultural factors on responses to assessment instruments. It seeks to identify cultural orientations that affect assessment instruments. Once cultural factors are identified, it will be possible to evaluate any measure as to the extent of bias on these factors and to attempt to control bias and increase validity of instruments. Therefore, existing instruments can still be used while controlling for the identified cultural factors.

The research is guided by several assumptions. First, Asian Americans may evaluate items on measures of psychopathology differently from Whites. These evaluations may be based on cultural factors such as social desirability, shame and stigma, familiarity with specific test items, defensiveness, conceptions of mental health, etc. Indeed, intra-Asian group differences may also exist. The task is to identify dimensions in which group differences are exhibited. Second, the validity of measures is threatened when evaluations significantly differ from one group to another. Cultural factors may suppress or enhance one's responses to assessment instruments. The task is to identify which cultural evaluations tend to influence responses to questionnaires. Third, once confounding cultural factors have been identified, it is possible to improve the validity of assessment instruments. The task is to make improvements in validity by "correcting" for bias or by constructing tests in which ethnic differences no longer exist on the identified dimensions. For example, let us assume that Asians are less likely than Whites to endorse a personality inventory item such as, "I have unusual sex practices." Let us also assume that Asians tend to give higher ratings of shame and stigma to the item. The ethnic differences in the endorsement of the item can be attributed to actual ethnic differences on the item *or* to differences on shame and stigma. Greater validity can be achieved by controlling for shame and stigma (by statistical means or by procedures similar to those used on the K-correction scale of the MMPI) or by constructing a test with items equally loaded for shame and stigma among different ethnic groups (similar to procedures on the Edwards Personal Preference Schedule [Edwards, 1959] in which respondents chose between items that are equated for social desirability).

The first study compared Asians and Whites in their performance on a measure of psychopathology, MMPI-2, in order to identify clinical scales in which group differences occur. As mentioned earlier, Asians reported more symptoms than Whites. In addition, less acculturated Asians reported more symptoms than more acculturated

Asians or Whites. The second study examines the influence of three hypothesized cultural dimensions on responses to the MMPI-2. Shame, symptom tolerance (i.e., whether the symptom is bothersome), and cultural familiarity (i.e., whether the symptom is common or frequent in the particular cultural group) were identified as important cultural dimensions for Asian Americans based on the past literature (e.g., Kim, 1978; Kitano, 1976; Sue & Morishima, 1982). The second study focused on whether ethnic differences in performance on the MMPI-2 can be explained by the cultural evaluations of the MMPI-2 items. Subjects are asked to rate the degree of shame, symptom tolerance, and cultural familiarity associated with each item of the MMPI-2. The data collection has been completed. In order to increase the generalizability of the findings, the data have been gathered from other universities across the U.S. as well as from UCLA. Once the important cultural dimensions are identified from the pilot studies, major studies will be proposed to develop scales that can control for the cultural biases.

### Assessment

The final point is addressed to practitioners. As noted earlier, skepticism has been voiced over assessment because of possible biases in the nosological systems; in the use of cognitive, personality, and psychopathology measures; and in making clinical inferences. Despite the skepticism, psychologists are frequently required to make evaluations in schools, mental health agencies, and courtrooms. What procedures can be used in such circumstances? Although issues of reliability and validity are involved, perhaps it is wise to distinguish two aspects, as noted in a previous paper of mine (Sue, 1988). The first deals with assessment procedures in general. The second includes special procedures that may be necessary with ethnic minority groups.

In any assessment task, the first step is to specify what one is interested in measuring (the referral question). The second step is to select the most appropriate inventory or test. Although factors such as the ease of administration, cost, degree of expertise required, etc., are often considered, reliability and validity of the measure for the characteristic of interest are the most important factors. Test manuals should include information on reliability and validity, as well as norms and samples upon which the norms are based. Of course, many assessment tools have not been adequately developed for different ethnic minority populations. With ethnic minority populations, there are some guidelines that are important to consider. These guidelines are not new. Nevertheless, they are important to reiterate.

1. Find tests that can be linguistically understood by clients. Also important is to determine the stimulus (linguistic) and conceptual equivalence of measures that are translated for the clients.
2. See if the test or assessment instrument has been standardized and normed on the particular ethnic minority group of the client. Increasingly, test developers are aware of the need to sample and validate tests and measures with different ethnic populations. For larger ethnic group populations, especially African Americans and Latino Americans, some measures have been standardized and normed. In the case of smaller populations, such as American Indians and Asian Americans, this is less likely to be the case.
3. If the test has not been standardized and normed on the group, exercise caution in interpreting the results. Tests and measures can still be useful, even if they have not been validated with a population. They provide samples of behaviors under standardized conditions. The primary issue is how to interpret findings. If the validity of a measure is uncertain, psychologists should exercise great care in interpreting the findings.
4. Test findings should be used to generate hypotheses for further testing. Although this is sound practice in general, this procedure is especially important in assessing members of ethnic minority groups because many assessment instruments may not have been validated with these groups.
5. Use multiple measures or multimethod procedures to see if tests provide convergent results. Before drawing conclusions, it is important to confirm findings from one instrument. This confirmation process should involve the administration of several different measures or different methods (e.g., behavioral ratings as well as self-reports) in order to see if the results are consistent.
6. Try to understand the cultural background of the client, in order to place test results in a proper context. Ethnic minority groups exhibit significant heterogeneity and individual differences. Individual differences exist in country of origin, language spoken and English proficiency, level of acculturation, ethnic identity, family structure, cultural values, history, etc. These differences have important implications for the ideal selection and interpretation of test results.
7. Enlist the aid of consultants who are familiar with the client's background and culture. It is difficult to know the cultures of



all the different ethnic groups in our society. Because cultural background has a major effect on assessment outcomes, the assistance of ethnic consultants is important. The consultants can help to place test findings in a proper cultural context.

Because of the growing multiethnic nature of our society and the increasing importance of assessment in all phases of life, there is an urgent need to direct our attention to the issues facing ethnic minority populations. A relatively small amount of research effort has been devoted to the valid assessment of these populations. The time is ripe for us to expend substantial efforts to address cross-cultural assessment issues.

## REFERENCES

- Adebimpe, V. (1981). Overview: White norms and psychiatric diagnosis of black patients. *American Journal of Psychiatry*, 138, 279-285.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- Banks, J. A. (1987). *Teaching strategies for ethnic studies*. Boston: Allyn and Bacon.
- Brislin, R. W. (1993). *Understanding culture's influence on behavior*. New York: Harcourt Brace Jovanovich.
- Chu, C., Lubin, B., & Sue, S. (1984). Reliability and validity of the Chinese Depression Adjective Checklist. *Journal of Clinical Psychology*, 40, 1409-1413.
- Cuellar, I., Harris, L., & Jasso, R. (1980). An acculturation scale for Mexican American normal and clinical populations. *Hispanic Journal of Behavioral Sciences*, 2, 199-217.
- Dohrenwend, B. P., & Dohrenwend, B. S. (1969). *Social status and psychological disorder*. New York: Wiley.
- Edwards, A. L. (1959). Edwards Personal Preference Schedule. San Antonio, TX: The Psychological Corporation.
- The forgotten: Chinese, lacking English, confined for 27 years. (1979, April 19). *Seattle Times*, p. A5.
- Garcia, J. (1981). The logic and limits of mental aptitude testing. *American Psychologist*, 36, 1172-1180.
- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Reviews of personality and social psychology* (vol. 2, pp. 141-165). Beverly Hills, CA: Sage.
- Good, B. J., & Good, M. D. (1986). The cultural context of diagnosis and therapy: A view from medical anthropology. In M. R. Miranda & H. H. L. Kitano (Eds.), *Mental health research & practice in*



*minority communities: Development of culturally sensitive programs* (pp. 1-27). Rockville, MD: National Institute of Mental Health.

Gordon, M. M. (1978) *Human nature, class, and ethnicity*. New York: Oxford University Press.

Hathaway, S. R., McKinley, J. C., & Butcher, J. N. (1989). *Minnesota Multiphasic Personality Inventory-2*. Minneapolis: University of Minnesota Press.

Helms, J. E. (1990). *Black and white racial identity*. Westport, CT: Greenwood Press.

Helms, J. E., & Carter, R. T. (1991). Relationships of White and Black racial identity attitudes and demographic similarity to counselor preferences. *Journal of Counseling Psychology*, 38, 446-457.

Hui, C. H. (1988). Measurement of individualism-collectivism. *Journal of Research in Personality*, 22, 17-36.

Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural counseling: A review and comparison of strategies. *Journal of Cross Cultural Psychology*, 16, 131-152.

Jones, E. E., & Thorne, A. (1987). Rediscovery of the subject: Intercultural approaches to clinical assessment. *Journal of Consulting and Clinical Psychology*, 55, 488-495.

Kaplan, R. M., & Saccuzzo, D. P. (1982). *Psychological testing: Principles, applications, and issues*. Monterey, CA: Brooks/Cole.

Keefe, K., Sue, S., Enomoto, K., Durvasula, R., & Chao, R. (in press). Asian American and White College Students' Performance on the MMPI-2. In J. N. Butcher (Ed.), *Handbook of international MMPI-2 research*. NY: Oxford University Press.

Kim, B. L. C. (1978). *The Asian-Americans: Changing patterns, changing needs*. Montclair, NJ: Association of Korean Christian Scholars in North America.

Kinzie, J. D., Manson, S. M., Vinh, D. T., Tolan, N. T., Anh, B., & Pho, R. N. (1982). Development and validation of a Vietnamese-language depression rating scale. *American Journal of Psychiatry*, 139, 1276-1281.

Kitano, H. H. L. (1976). *Japanese-Americans: The evaluation of a subculture*. Englewood Cliffs, NJ: Prentice-Hall.

Kleinman, A. M. (1977). Depression, somatization, and the new cross-cultural psychiatry. *Social Science and Medicine*, 11, 3-10.

Kleinman, A. (1991, April). *Culture and DSM-IV: Recommendations for the introduction and for the overall structure*. Paper presented at the Conference on Culture and DSM-IV, Pittsburgh.

Korchin, S. J. (1980). Clinical psychology and minority problems. *American Psychologist*, 35, 262-269.

LaFromboise, T. D. (1988). American Indian mental health policy. *American Psychologist*, 43, 388-397.

Li-Repac, D. (1980). Cultural influences on clinical perception: A comparison between Caucasian and Chinese-American therapists. *Journal of Cross-Cultural Psychology*, 11, 327-342.

Lopez, S. R. (1989). Patient variable biases in clinical judgment: Conceptual overview and methodological considerations. *Psychological Bulletin*, 106, 184-204.

Mendoza, R. H. (1989). An empirical scale to measure type and degree of acculturation in Mexican-American adolescents and adults. *Journal of Cross-Cultural Psychology*, 20, 372-385.

Mercer, J. R., & Lewis, J. R. (1979). *System of Multi-Cultural Pluralistic Assessment: Conceptual and technical manual*. San Antonio, TX: The Psychological Corporation.

Mukherjee, S., Shukla, S., Woodle, J., Rosen, A. M., & Olarte, S. (1983). Misdiagnosis of schizophrenia in bipolar patients: A multiethnic comparison. *American Journal of Psychiatry*, 140, 1571-1574.

Neighbors, H. W., Jackson, J. S., Campbell, L., & Williams, D. (1989). The influence of racial factors on psychiatric diagnosis: A review and suggestions for research. *Community Mental Health Journal*, 25(4), 301-311.

Ottavi, T. M., Pope-Davis, D. B., & Dings, J. G. (1994). Relationship between white racial identity attitudes and self-reported multicultural counseling competencies. *Journal of Counseling Psychology*, 41, 149-154.

Padilla, A. M., & Salgado DeSnyder, N. (1985). Counseling Hispanics: Strategies for effective intervention. In P. Pedersen (Ed.), *Handbook of cross-cultural counseling and therapy* (pp. 157-164). Westport, CT: Greenwood Press.

Phinney, J. S. (1992). The multigroup ethnic identity measure: A new scale for use with diverse groups. *Journal of Adolescent Research*, 7, 156-176.

Ponterotto, J. G., & Casas, J. M. (1991). *Handbook of racial/ethnic minority counseling research*. Springfield, IL: Charles C. Thomas.

Rogler, L. H., Malgady, R. G., & Rodriguez, O. (1989). *Hispanics and mental health: A framework for research*. Malabar, FL: Krieger Publishing Company.

Sodowsky, G. R., & Plake, B. S. (1991). Psychometric properties of the American-International Relations Scale. *Educational and Psychological Measurement*, 51, 207-216.

Sodowsky, G. R., Taffe, R. C., Gutkin, T. B., & Wise, S. L. (1994). Development of the Multicultural Counseling Inventory: A self-

report measure of multicultural competencies. *Journal of Counseling Psychology*, 41, 137-148.

Spitzer, R. L., Gibbon, M., & Endicott, J. (1985). *Global Assessment Scale*. New York: Department of Research Assessment and Training, New York State Psychiatric Institute.

Sue, D., Ino, S., & Sue, D. M. (1983). Nonassertiveness of Asian Americans: An inaccurate assumption? *Journal of Counseling Psychology*, 30, 581-588.

Sue, D., & Sue, S. (1987). Cultural factors in the clinical assessment of Asian Americans. *Journal of Consulting and Clinical Psychology*, 55(4) 479-487.

Sue, D. W., & Sue, D. (1991). *Counseling the culturally different: Theory and practice*. New York: Wiley.

Sue, S. (1988). *Sociocultural issues in the assessment and classroom teaching of language minority students*. Crosscultural Special Education Series, V.3. Sacramento: California State Department of Education.

Sue, S., & Abe, J. (1988). *Predictors of academic achievement among Asian American and White students*. New York: The College Board.

Sue, S., & Morishima, J. (1982). *The mental health of Asian Americans*. San Francisco: Jossey-Bass Publishers.

Sue, S., Wagner, N. N., Ja, D., Margullis, C., & Lew, L. (1976). Conceptions of mental illness among Asian and Caucasian American students. *Psychological Reports*, 38, 703-708.

Suinn, R., Rickard-Figueroa, K., Lew, S., & Vigil, P. (1987). The Suinn-Lew Asian Self-Identity Acculturation Scale: An initial report. *Educational and Psychological Measurement*, 47, 401-407.

Takeuchi, D., & Speechley, K. N. (1989). Ethnic differences in the marital status and psychological distress relationship. *Social Psychiatry and Psychiatric Epidemiology*, 24, 288-294.

Westermeyer, J. (1987). Cultural factors in clinical assessment. *Journal of Consulting and Clinical Psychology*, 55(4), 471-478.

White, J. L. (1984). *The psychology of blacks: An Afro-American perspective*. Englewood Cliffs, NJ: Prentice Hall.

Williams, R. L. (1974). Scientific racism and IQ: The silent mugging of the Black community. *Psychology Today*, 7, 32-41.

Yang, K. S., & Bond, M. H. (1990). Exploring implicit personality theories with indigenous or imported constructs: The Chinese case. *Journal of Personality and Social Psychology*, 58, 1087-1095.

Zane, N., & Sue, S. (1986). Reappraisal of ethnic minority issues: Research alternatives. In E. Seidman & J. Rappaport (Eds.), *Redefining social problems* (pp. 289-304). New York: Plenum.

Zane, N. (1991, August). *An empirical examination of loss of face among Asian Americans*. Paper presented at the annual meeting of the American Psychological Association, San Francisco.